TravelBot: Journey Disruption Alerts Utilising Social Media and Linked Data⋆

David Corsar, Milan Markovic, Paul Gault, Mujtaba Mehdi, Peter Edwards,
John D. Nelson, Caitlin Cottrill, and Somayajulu Sripada

dot.rural Digital Economy Hub, University of Aberdeen, Aberdeen, AB24 5UA
{dcorsar,m.markovic,p.gault,mmehdi,p.edwards,
j.d.nelson,c.cottrill,yaji.sripada,}@abdn.ac.uk

Abstract. This demo paper presents a travel advice system based on information extracted from social media and linked data.

Keywords: Social Media, Twitter, Transport Disruption, Linked Data

1 Introduction

The Twitter[1] microblogging platform is widely used in the public transport domain by passengers to communicate with transport operators and by operators to provide customer service and passenger information [1,2]. In particular, the reliable, low-cost information distribution provided by Twitter has made it an important channel for publishing real-time updates about disruptions to the transport network and services [2]. However, to benefit from this passengers must first find such Tweet(s), which can be published by any Twitter user including transport operators, relevant authorities, local media outlets, and other passengers. Travellers must then have the necessary knowledge to evaluate the quality of the information conveyed in terms of its veracity, temporal and geospatial relevance to their journey, and reliability of the provider. Finally, they must decide if the disruption will adversely impact their journey and, if so, whether any changes to their travel plans are necessary.

This demo will show the *TravelBot* system developed to support bus users in the city of Aberdeen, UK. The demo will feature: a user registering a journey with the system; *TravelBot* monitoring Twitter for messages describing transport related events that may disrupt that journey; and when one is detected, sending a personalised message to the user warning them of the potential disruption[2]. The demo will utilise the datasets and system shown in Fig. 1.

---

[1] http://twitter.com/
[2] A video of this demo is available at https://youtu.be/ZAg6RnCQoUI.

2 The TravelBot Ecosystem

The *TravelBot* system[3] is supported by a linked transport information ecosystem (illustrated in Fig. 1 and further discussed below) that is based on a series of ontologies. Services provide the system functionalities by reasoning with data accessed via SPARQL endpoints. Figure 2 presents a sample of the data generated for a Tweet, a user journey, and an alert message sent to a user.
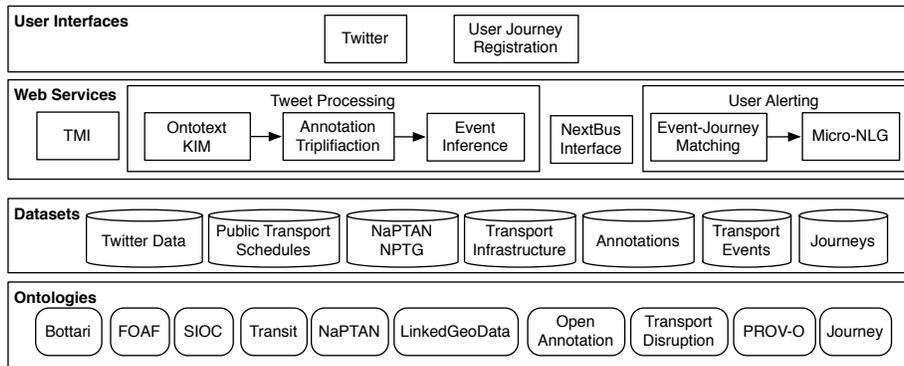

Fig. 1. The ecosystem supporting the *TravelBot* system.

Tweets published by accounts known to provide travel information for the geographic area, including bus operators, transport authorities and registered users, along with Direct (private) Messages to the *TravelBot* Twitter account are received and stored by the Twitter Monitoring Infrastructure[4] (TMI). The message and associated metadata (including its unique identifier, author, and creation timestamp) are stored in the Twitter Data dataset and published as linked data using the Bottari[5], FOAF[6], and SIOC[7] ontologies[8].

Once stored, the message's URI is passed to the Tweet Processing component, which extracts, classifies, and contextualises a semantic representation of any transport event(s) described in the message. The Ontotext KIM[9] platform, which is designed to identify semantic entities in text, is configured to discover entities related to transport events in the message. To achieve this, the KIM knowledge base has been extended with RDF descriptions, including names,

---

[3] The TravelBot system is available at https://github.com/SocialJourneys.

[4] The TMI system is available at https://github.com/SocialJourneys/TMI.

[5] http://purl.org/NET/bottari.n3

[6] http://xmlns.com/foaf/0.1/

[7] http://rdfs.org/sioc/ns#

[8] To comply with Twitter's terms and conditions, this data is only available within this system.

[9] http://ontotext.com/kim

commonly used abbreviations and slang terms describing: types of potentially disruptive transport events related to network operator actions, public transport, and traffic described by the Transport Disruption ontology[10]; open bus service and schedule data[11] stored in the Public Transport Schedules dataset[12]; public transport access points from the NaPTAN dataset[13] and settlements from the NPTG dataset[14]; and the road network extracted from openstreetmap.org and stored in the Transport Infrastructure dataset[15].

The Annotation Triplification component generates annotations for each entity identified by KIM, represented with the Open Annotation ontology[16]. As shown in Fig. 2, each annotation links to the source message and the identified resource; these are stored in the Annotations dataset. The Event Inference module uses a set of hand crafted SPIN[17] rules to create a semantic representation of transport events based on these annotations and represented using the Transport Disruption ontology. The rules attempt to determine if a transport related event is described, and if so, to associate a geolocation and time period with it, attribute values specific to the event type (for example, in Fig. 2 the carriageway affected by the roadworks) and link to any bus services that may be affected. The inferred event(s) is (are) added to the Transport Events dataset, along with provenance information[18] recording the creation timestamp and the message that it was derived from.

TravelBot users register any journey for which they wish to receive notifications. Each journey is described in terms of the week days on which it is made, the time of travel, bus service(s) used, and boarding and alighting locations. At a user specified time before each journey, the NextBus Interface component retrieves the upcoming arrival times for bus(es) on the specified service(s) at the boarding bus stop[19]. This is privately communicated to the user via Twitter as a Direct Message and will be available through their usual Twitter client.

The Event Journey Matching component uses a series of quality metrics to determine if any events might disrupt a user's journey. The metrics include: temporal relevance, which considers if the event is ongoing during the user's journey; geospatial relevance, which considers if the event's location and user's

---

[10] http://purl.org/td/transportdisruption#

[11] Published at http://www.travelinedata.org.uk/traveline-open-data/traveline-national-dataset/.

[12] Represented using the Transit ontology - http://vocab.org/transit/terms/.

[13] http://data.gov.uk/dataset/naptan, represented using the NaPTAN ontology - http://transport.data.gov.uk/def/naptan/.

[14] http://data.gov.uk/dataset/nptg, also represented with the NaPTAN ontology.

[15] Represented using the LinkedGeoData ontology - http://linkedgeodata.org/ontology/.

[16] http://www.w3.org/ns/oa#

[17] http://spinrdf.org/

[18] Represented using PROV-O - http://www.w3.org/ns/prov#.

[19] This uses the NextBus API (http://www.travelinedata.org.uk/traveline-open-data/nextbuses-api/), which provides real-time arrival information for all bus stops in the UK.
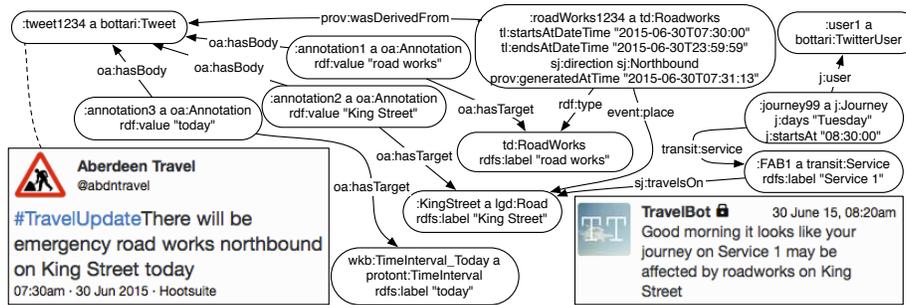
Fig. 2. Sample annotations for a Tweet published by a trusted source, the inferred event description (:roadworks1234), a user journey (:journey99), and message sent to the user warning of potential disruption to a journey.

expected route of travel overlap, with more precise event locations (e.g. a road the bus travels along) being assigned a higher relevance than less precise locations (e.g. a locality the bus travels through); service relevance, which considers if the event is known to affect a bus service that the user will travel on; and veracity, which is based on if the Tweet's author is included in a predefined set of trusted users (e.g. the bus operator, transport authorities, or a local radio station)[20].

If an event is rated with sufficiently high temporal relevance and either geospatial or service relevance, then a tailored message is generated by the Micro-Natural Language Generation (Micro-NLG) component, and sent to the user. The message is based on the inferred event resource and attempts to convey the level of certainty that the journey will be affected, as indicated by the quality metrics. For example, in Fig. 2 although the roadworks are on a road that the bus travels on, no delays have been reported by the operator so the message is deliberately vague using the phrase "may be affected"; if the operator does report delays on Service 1 in that area, this would change to a stronger phrase such as "is highly likely to be delayed".

A user study is planned to evaluate the *TravelBot* user experience and system performance with inferring event descriptions from social media posts.

References

1. T. Camacho, M. Foth, and A. Rakotonirainy. Pervasive technology and public transport: Opportunities beyond telematics. *Pervasive Computing, IEEE*, 12(1):18–25, Jan 2013.
2. P. Gault, D. Corsar, P. Edwards, J. D. Nelson, and C. Cottrill. You'll Never Ride Alone: The Role of Social Media in Supporting the Bus Passenger Experience. In *Ethnographic Praxis in Industry Conference Proceedings*, volume 2014, pages 199–212, 2014.

---

[20] While this can be considered a basic metric for event veracity, metrics considering other factors, such as the number of reports about an event, could be developed.